REMARKS

Claims 61, 66, and 72 have been cancelled. Claims 1, 2, 3, 15, 20-22, 34, 39-41, 60, 62, 65, 67, 70, 71, and 73 have been amended. New claims 74-79 have been added. Claims 1-4, 9, 12, 15, 20-23, 28, 31, 34, 39-42, 47, 50, 58-73, and 74-79 are pending in the instant application.

It is respectfully submitted that the present amendment presents no new issues or new matter and places this case in condition for allowance.

I. The Rejection of Claims 1-4, 9, 12, 15, 20-23, 28, 31, 34, 39-42, 47, and 50-73 under 35 U.S.C. § 112, Second Paragraph

Claims 1-4, 9, 12, 15, 20-23, 28, 31, 34, 39-42, 47, and 50-73 stand rejected under 35 U.S.C. § 112, second paragraph, as being indefinite on the following grounds.

Ground 1: The Office Action states that claim 1 is allegedly vague and indefinite because it is unclear whether the first and second nucleic acid sequences are a fusion/hybrid sequence. This rejection is respectfully traversed.

Applicants respectfully point out that claim 1 recites in part "the mutant cell comprises a first nucleic acid sequence encoding the heterologous protein and a second nucleic acid sequence comprising a mutation of at least one of the genes *cypX* and *yvmC*". One of ordinary skill in the art would readily understand that the "first nucleic acid sequence" is separate from the "second nucleic acid sequence". If the "first nucleic acid sequence" was fused to the "second nucleic acid sequence", the claim would use the term "fused to" or "linked in frame" in place of the word "and" in the wording above. Applicants submit that the claim language is clear and definite, but would consider any suggestion by the Office to reword the claim language.

Ground 2: The Office Action states that claims 15, 34, and 70 are allegedly vague and indefinite because they require that the mutant cell produce at least about 25% less of the red pigment compared to the parent *Bacillus* cell when cultured under identical conditions and it is unclear what type of mutations would render a cell with this phenotype. This rejection is respectfully traversed.

The Office states that it is unclear what type of mutations would render a cell with this phenotype. The term "modification" is defined on page 3, lines 27-31, of the specification as an introduction, substitution, or removal of one or more nucleotides in the gene or a regulatory element required for the transcription or translation thereof; a gene disruption; gene conversion; a gene deletion; or random or specific mutagenesis of at least one of the genes *cypX* and *yvmC*." One type of modification that could reduce the amount of red pigment produced is to

substitute the promoter associated with the *cypX* and *yvmC* gene with a weaker promoter or to mutate the promoter by introducing, substituting, or removing one or more nucleotides in the promoter so less red pigment is produced. However, solely to further prosecution, Applicants have amended claims 15, 34, and 70 to recite in part "the mutant cell produces no detectable red pigment compared to the parent *Bacillus* cell when cultured under identical conditions".

Ground 3: The Office Action states that claims 60, 65, and 71 are vague and indefinite due to the phrase "comprises a mutation of one or more genes which encode a protease" because the claim fails to teach what type of mutation or what phenotype is rendered by the mutation and what genes are encompassed by the term "protease". This rejection is respectfully traversed.

Preliminarily, claims 60, 65, and 71 have been amended to recite in part: "the mutant cell is further deficient in the production of protease".

The Office argues that the term "protease" is used for many other proteins/protein fragments in the art. Applicants disagree. The term "protease" simply means a group of enzymes that catalyze the hydrolysis of the peptide bond in peptides or proteins. Applicants provide a copy of the definition for proteases fro the Concise Encyclopedia Biochemistry, Second Edition (Attachment 1). Applicants submit that the term "protease" is clear and definite to one of ordinary skill in the art. Moreover, Applicants disclose on page 8, lines 24-28, of the specification examples of protease genes (*aprE* and *nprE*) that can be inactivated to render a *Bacillus* cell deficient in protease.

Ground 4: The Office Action states that in claims 60, 61, 62, 66, 72 and 73 are vague and indefinite due to the phrase a "mutant cell [which] further comprises a modification of one or more genes selected from the group consisting of *spollAC*, *srfA*, *srfB*, *srfC*, *srfD*, and *amyE*" because it is unclear what is encompassed by the term "modification" and the mere recitation of a name, *i.e.*, *spollAC*, *srfA*, *srfB*, *srfC*, *srfD*, *amyE*, *nprE*, *aprE*, etc., to describe the invention is not sufficient to satisfy the Statute's requirement of adequately describing and setting forth the inventive concept. This rejection is respectfully traversed.

Preliminarily, claims 60, 65, and 71 have been amended to recite in part: "the mutant cell is further deficient in the production of protease"; and claims 62, 72, and 73 have been amended to recite in part: "the mutant cell is further deficient in the production of surfactin."

First, the Office states that it is unclear what is encompassed by the term "modification". The term "modification" is defined on page 3, lines 27-31, of the specification.

Moreover, the Office states that "the mere recitation of a name, *i.e.*, *spollAC*, *srfA*, *srfB*, *srfC*, *srfD*, *amyE*, *nprE*, *aprE*, etc., to describe the invention is not sufficient to satisfy the

Statute's requirement of adequately describing and setting forth the inventive concept". As described above, the claims have been amended. It is well settled that a patent need not teach and preferably omits what is well known in the art. *Spectra-Physics Inc. v. Coherent Inc.*, 3 U.S.P.Q.2d 1737 (Fed. Cir. 1987). For example, a *Bacillus* cell deficient in the production of surfactin is described in U.S. Patent No. 5,958,728 (attached), which is mentioned on page 8, line 32, of the specification.

For the foregoing reasons, Applicants submit that the claims overcome the rejections under 35 U.S.C. § 112 and respectfully request reconsideration and withdrawal of the rejection.

II. The Rejection of Claims 1-4, 9, 12, 15, 20-23, 28, 31, 34, 39-42, 47, and 50-73 under 35 U.S.C. § 112, First Paragraph

Claims 1-4, 9, 12, 15, 20-23, 28, 31, 34, 39-42, 47, and 50-73 stand rejected under 35 U.S.C. § 112, first paragraph, as failing to comply with the enablement requirement. The Office Action stated:

> The specification only teaches the *cypX* gene set forth in SEQ ID NO:1 and the *yvmC* gene set forth in SEQ ID NO:7 from *Bacillus subtilis*. It is unclear and unpredictable whether the other 14 species of *Bacillus* recited in the instant claims, or the more than **208** species of Bacillus currently known and categorized, possess *yvmC* and/or *cypX* genes. At the very least, the specification has only shown that *B. licheniformis* possesses a *cypX* gene which can produce red pigment. The prior art is silent as to whether any other species of Bacillus possess the *cypX* and *yvmC* proteins and, therefore, it would take one of skill in the art undue experimentation in order to isolate the claimed DNA sequences from any species of bacteria other than *B. subtilis*. The specification only enables deleting or mutating the *cypX* and *yvmC* from B. subtilis and the *cypX* gene from *B. licheniformis* in order to get better expression of a heterologous protein.

This rejection is respectfully traversed for the reasons of record and further for the reasons stated below.

First, Applicants disclose in Example 5 methods for identifying *Bacillus* strains that produce a red pigment during fermentation.

Second, the specification contains an extensive disclosure of techniques that are well known in the art and indeed routine for persons of ordinary skill in the art for identifying and using other *cypX* and *yvmC* genes. Applicants describe methods for identifying *cypX-yvmC* operons using DNA microarrays (Example 1); for isolating *cypX-yvmC* operons from *Bacillus subtilis* and *Bacillus licheniformis* (Examples 2 and 6); for constructing *Bacillus subtilis* strains and *Bacillus licheniformis* strains with disrupted *cypX-yvmC* operons (Examples 2-4, Example 6; and page 4, line 24, to page 7, line 10, of the specification); for fermenting *Bacillus* strains to

show the absence of red pigment formation; for determining the degree of homology between two nucleic acid sequences using the Wilbur-Lipman method according to Wilbur and Lipman, 1983, *Proceedings of the National Academy of Science USA* 80: 726-730 (page 12, line 29, to page 13, line 8); and for expressing heterologous biological substances in red pigment-deficient *Bacillus* strains (page 11, lines 33, to page 17, line 19). It is well within the skill of the art to discover new *cypX* and *yvmC* genes in other species of *Bacillus* using Applicants' disclosure, as exemplified in Example 6 of the specification for *Bacillus licheniformis*. Applicants clearly demonstrate that it would NOT require undue experimentation to discover new *cypX* and *yvmC* genes in any of the other 14 species of *Bacillus*.

Third, preliminarily, Applicants would like to correct a statement in the Amendment of December 1, 2006, which read: "Applicants have shown in Example 6 that primers based on the *cypX* gene from *Bacillus subtilis* were used to clone by PCR the *cypX* gene from *Bacillus licheniformis* and delete a portion of the gene to prevent formation of the red pigment." Applicants meant to state the following: "Primers based on the sequence of the *cypX* gene from *Bacillus subtilis* can be used to design degenerate primers for cloning by PCR the *cypX* gene from *Bacillus licheniformis*". Consequently, one of ordinary skill in the art can use the *cypX* and *yvmC* genes isolated from *Bacillus subtilis* (see Examples 1 and 2) to synthesize a series of degenerate primers to clone by PCR the corresponding genes from other *Bacillus* species that are involved in the production of the red pigment during fermentation and delete a portion of such genes using Applicants' specification to prevent formation of the red pigment.

An alignment of other *cypX* genes to the *cypX* gene disclosed in the instant application or other *yvmC* genes to the *yvmC* gene also disclosed in the instant application was conducted by Applicants against the Uniref100 (Uniprot) protein database using the BLASTX algorithm. The results (Attachment 2 for *cypX* and Attachment 3 for *yvmC*) showed that in addition to the *Bacillus subtilis* genes, similar genes are present in *Bacillus licheniformis* ATCC 14580 and *Bacillus thuringiensis* ATCC 35646. The percent identity of the deduced amino acid sequences of the *Bacillus licheniformis* and *Bacillus thuringiensis cypX* genes compared to the *Bacillus subtilis cypX* gene were 89% and 72%, respectively. The percent identity of the deduced amino acid sequences of the *Bacillus licheniformis* and *Bacillus thuringiensis yvmC* genes compared to the *Bacillus subtilis yvmC* gene were 70% and 57%, respectively. Consequently, as Applicants stated above, one of ordinary skill in the art can use the *cypX* and *yvmC* genes isolated from *Bacillus subtilis* (see Examples 1 and 2) to synthesize a series of degenerate primers based on either the *cypX* gene or *yvmC* gene from *Bacillus subtilis* to clone by PCR corresponding genes from other *Bacillus* species.

Fourth, the Office provides that Genentech Inc. v. Novo Nordisk A/S (CAFC) 42 USPQ2d 1001 clearly states: "Patent protection is granted in return for an enabling disclosure of an invention, not for vague intimations of general ideas that may or may not be workable. See Brenner v. Manson, 383 U.S. 519, 536, 148 USPQ 689, 696 (1966) (stating, in context of the utility requirement, that "a patent is not a hunting license. It is not a reward for the search, but compensation for its successful conclusion.") Tossing out the mere germ of an idea does not constitute enabling disclosure. While every aspect of a generic claim certainly need not have been carried out by an inventor, or exemplified in the specification, reasonable detail must be provided in order to enable members of the public to understand and carry out the invention."

Applicants disagree with the Office's assertion that Genentech Inc. v. Novo Nordisk A/S applies to the instant application. Applicants respectfully disagree that they have tossed out a mere germ of an idea.

The Genentech Inc. v. Novo Nordisk A/S court stated:

> The question before us is whether the specification would have enabled a person having ordinary skill in the art at the time of filing to use cleavable fusion expression to make hGH without undue experimentation. There is no dispute that the portion of the specification chiefly relied upon by Genentech and by the district court, column 7, lines 29-59, does not describe in any detail whatsoever how to make hGH using cleavable. fusion expression. For example, no reaction conditions for the steps needed to produce hGH are provided; no description of any specific cleavable conjugate protein appears. The relevant portion of the specification merely describes three (or perhaps four) applications for which cleavable fusion expression is generally well-suited and then names an enzyme that might be used as a cleavage agent (trypsin), along with sites at which it cleaves ("arg-arg or lys-lys, etc."). Thus, the specification does not describe a specific material to be cleaved or any reaction conditions under which cleavable fusion expression would work.
>
> ...
>
> It is true, as Genentech argues, that a specification need not disclose what is well known in the art. See, e.g., Hybritech Inc. v. Monoclonal Antibodies, Inc., 802 F.2d 1367, 1385, 231 USPQ 81, 94 (Fed. Cir. 1986). However, that general, oft-repeated statement is merely a rule of supplementation, not a substitute for a basic enabling disclosure. It means that the omission of minor details does not cause a specification to fail to meet the enablement requirement. However, when there is no disclosure of any specific starting material or of any of the conditions under which a process can be carried out, undue experimentation is required; there is a failure to meet the enablement requirement that cannot be rectified by asserting that all the disclosure related to the process is within the skill of the art. It is the specification, not the knowledge of one skilled in the art, that must supply the novel aspects of an invention in order to constitute adequate enablement. This specification provides only a starting point, a direction for further research.

As the court above stated, the specification chiefly relied upon by Genentech did not describe in any detail whatsoever how to make hGH using cleavable fusion expression. For example, no reaction conditions for the steps needed to produce hGH were provided; no description of any specific cleavable conjugate protein appeared. Applicants submit that the

specification described in Genentech Inc. v. Novo Nordisk A/S is in stark contrast to the reasonable and extensive detail provided in Applicants' specification. Applicants' specification discloses methods for identifying *cypX-yvmC* operons; for isolating *cypX-yvmC* operons from *Bacillus subtilis* and *Bacillus licheniformis*; for constructing *Bacillus subtilis* strains and *Bacillus licheniformis* strains with disrupted *cypX-yvmC* operons; for fermenting *Bacillus* strains to show the absence of red pigment formation; and for expressing heterologous biological substances in red pigment-deficient *Bacillus* strains. Consequently, Applicants have not tossed out a mere germ of an idea.

Fifth, if one skilled in this art wished to identify and isolate a *cypX* or *yvmC* gene other than the one disclosed in Applicants' specification, she/he can merely read Applicants' specification for directions how to identify and isolate such a gene. Applicants' claimed invention is not complicated, and no special equipment or unusual procedures must be provided when practicing the invention. The methods for cloning another gene and determining its relatedness to the disclosed genes is well described in Applicant's specification and well known in the art. One skilled in the art would then merely have to substitute the other gene. Thus, there is no basis for concluding that persons skilled in this art, armed with the specification, would have to exercise undue experimentation to determine which genes to use in Applicants' claimed invention and which ones not to use. While some experimentation is necessary to identify other genes, such experimentation is a simple process and well known in the art. The experimentation would not be undue and certainly would not require ingenuity beyond that expected of one of ordinary skill in the art.

Applicants submit, therefore, that the claimed inventions are enabled by the specification based on the specification's extensive disclosure of techniques which are well known in the art and indeed routine for persons of ordinary skill for making and using the claimed subject matter of the present invention. Based on Applicants' disclosure, it would be routine for one of ordinary skill in the art to make and use the claimed nucleic acid sequences.

For the foregoing reasons, Applicants submit that the rejections under 35 U.S.C. § 112 have been overcome and respectfully request reconsideration and withdrawal of the rejection.

III. The Rejection of Claims 1-4, 9, 12, 15, 20-23, 28, 31, 34, 39-42, 47, and 50-73 under 35 U.S.C. § 112, First Paragraph

Claims 1-4, 9, 12, 15, 20-23, 28, 31, 34, 39-42, 47, and 50 stand rejected under 35 U.S.C. § 112, first paragraph, as failing to comply with the written description requirement. The Office Action stated:

> The specification only teaches the *cypX* gene set forth in SEQ ID NO: 1 and the *yvmC* gene set forth in SEQ ID NO: 7 from *Bacillus subtilis*. It is unclear and unpredictable whether the other 14 species of *Bacillus* recited in claims 12, 31 and 50 possess red pigment genes, much less red pigment genes with the sequences set forth in SEQ ID NOs: 1 and 7. The specification only provides adequate written description for methods which use *B. subtilis* genes and mutations of the *cypX* gene set forth in SEQ ID NO: 1 and the *yvmC* gene set forth in SEQ ID NO: 7 and not the broad scope of the claims.

This rejection is respectfully traversed for the reasons of record and further for the reasons below.

The Office argues that Applicants have not identified any common structural core which one skilled in the art could use to identify any genus of polynucleotides and Applicants are claiming such polynucleotide homologues only by their functionality, that of producing red pigment. Applicants disagree for the reasons stated below.

Federal Circuit decisions provide that a specification complies with the written description requirement if it provides "a precise definition, such as by structure, formula, chemical name, or physical properties of the claimed subject matter sufficient to distinguish it from other materials." See, e.g., *University of California v. Eli Lilly and Co.*, 43 U.S.P.Q.2d 1398, 1404 (Fed. Cir. 1997); *Enzo Biochem v. Gen-Probe Inc.*, 63 U.S.P.Q.2d 1609, 1613 (Fed. Cir. 2002). In fact, "[a] description of a genus of cDNAs may be achieved by means of a recitation of a representative number of cDNAs, defined by nucleotide sequence, falling within the scope of the genus or of a recitation of structural features common to the members of the genus, which features constitute a substantial portion of the genus." *Eli Lilly and Co.*, 43 U.S.P.Q.2d at 1569. The test is not whether one of ordinary skill in the art envisions all of the claimed subject matter, as suggested in the Office Action.

In the Amendment of December 1, 2006, Applicants submitted that it is well established in the art that the definition of a genus of genes is accomplished by using structural features that show the relatedness of the genes and their encoded products and predict the function of polypeptides encoded by novel genes. Such structural features are (1) percent identity of the amino acid sequences encoded by the genes, (2) percent homology of the nucleic acid sequences of the genes, and (3) nucleic acid hybridizations under defined stringent conditions to identify complementary strands of genes encoding the same or similar enzyme or protein function. In the claims at issue, Applicants provide a recitation of one of these structural features common to the claimed genera: (1) a nucleic acid sequence having at least 70% homology to SEQ ID NO: 1, and (2) a nucleic acid sequence having at least 70% homology to SEQ ID NO: 7.

Percent homology is highly predictive of function and without this tool it would be

impossible to make meaningful annotations of genomes in sequencing projects. Genes that share 70% homology are known to encode products that possess the same catalytic/biochemical function which has formed the basis for genome annotation and comparative genomics. A simple search of any public database using the criteria above for a reference gene of interest will prove that there is a definitive relationship between function and % homology at either the nucleotide or amino acid level.

Since the claimed structural feature of percent homology specifies a family of structurally- and functionally-related genes and provides a correlation between function and structure, Applicants have identified a structural core which one skilled in the art could use to identify other *cypX* and *yvmC* genes from other Bacilli. To solely further prosecution, however, the claims have been amended to recite "95% homology".

The Office also argues that the specification does not provide evidence that one skilled in the art would know what modifications, and what regions of the *yvmC* or *cypX* coding regions to target for modifications, in order to produce the desired bacterium. Applicants disagree. Applicants have provided a number of alternatives to produce a mutation that inactivates at least one of the genes *cypX* and *yvmC*. These alternatives are described on page 4, line 24, to page 7, line 10, of the specification and include gene deletion; introducing, substituting, or removing one or more nucleotides in the gene; gene disruption; gene conversion; anti-sense techniques; and random or specific mutagenesis. However, one skilled in the art would recognize that the most efficient method for inactivating such genes would involve gene deletion or gene disruption. Applicants used gene deletion as described in Examples 3, 4, and 6. Consequently, the Office's statement that one skilled in the art would not know what modifications and what regions of the *yvmC* or *cypX* coding regions to target to inactivate such a gene is without foundation.

The Office cites Bowie *et al.*, *Science* 247:1306-1310, page 1306, which presents a discussion on the tolerance of proteins to amino acid substitutions, and the Office argues that although the reference is a discussion of protein substitutions, as the present case is concerned with polynucleotides encoding such proteins, the teachings of the reference are equally applicable to the mutations of the claimed inventions on the ground that one skilled in the art would not be able to recognize from the current disclosure any substitutions, or other mutation (except, perhaps, deletion of the whole polynucleotide) that would result in a decreased gene product activity. As Applicants state above, the specification provides many alternatives for inactivating a *cypX* or *yvmC* gene and exemplify one of the most efficient methods. Applicants are not required to exemplify every method it discloses. Moreover, the Bowie *et al.* reference is

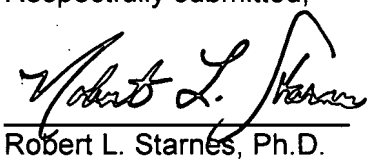
more pertinent to the construction of variants of a particular protein and one of ordinary skill in the art would not be inclined to use, for example, point mutations to inactivate a gene, when there are more efficient methods as used by Applicants.

For the foregoing reasons, Applicants submit that the new claims overcome the rejections under 35 U.S.C. § 112 and respectfully request reconsideration and withdrawal of the rejection.

IV. Conclusion

In view of the above, it is respectfully submitted that all claims are in condition for allowance. Early action to that end is respectfully requested. The Examiner is hereby invited to contact the undersigned by telephone if there are any questions concerning this amendment or application.

Respectfully submitted,



Date: August 22, 2007

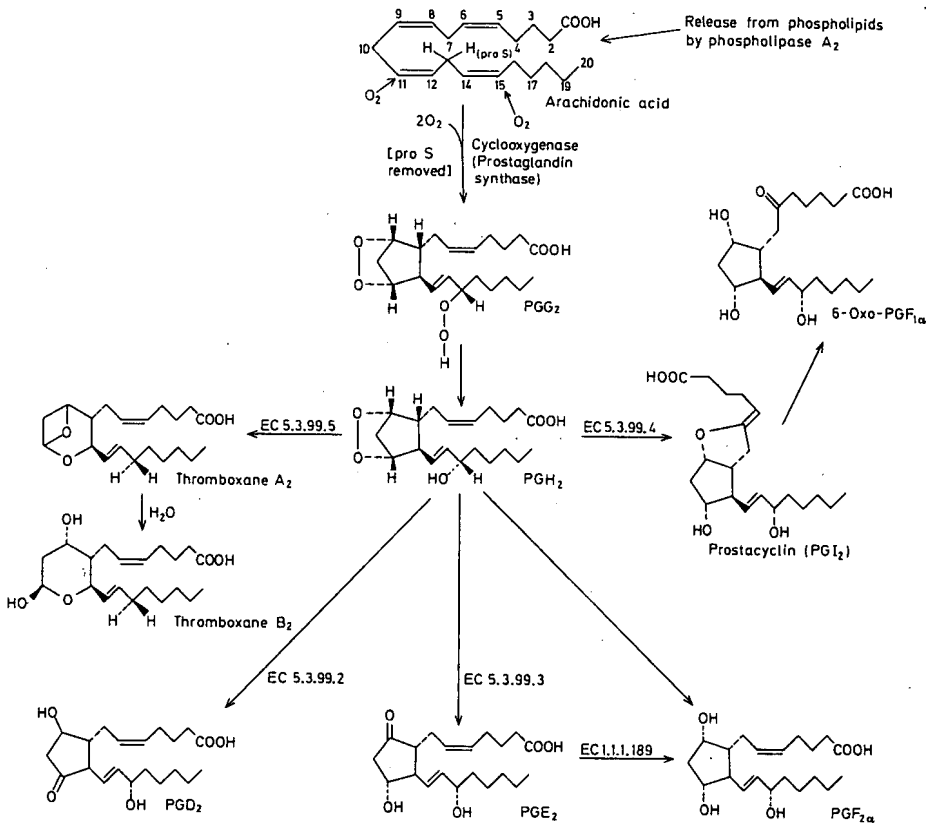
Robert L. Starnes, Ph.D.
Reg. No. 41,324
Novozymes, Inc.
1445 Drew Avenue
Davis, CA 95618
(530) 757-8100

Concise Encyclopedia Biochemistry

Second Edition

Revised and expanded by
Thomas Scott and Mary Eagleson

Walter de Gruyter
Berlin · New York 1988



The cyclooxygenase pathway of arachidonate metabolism.
EC 1.1.1.189: Prostaglandin-E_2 9-oxoreductase
EC 5.3.99.2: Prostaglandin-H_2D-isomerase
EC 5.3.99.3: Prostaglandin-H_2E-isomerase
EC 5.3.99.4: Prostacyclin synthase
EC 5.3.99.5: Thromboxane synthase

sis of thromboxanes at lower concentrations than are required to inhibit PG synthesis; thus low concentrations of aspirin enhance bleeding by shifting the metabolism of PGH_2 toward PGI_2 and away from thromboxanes. [S. Moncada & J.R. Vane, "Pharmacology and endogenous roles of Prostaglandin Endoperoxides, Thromboxane A_2 and Prostacyclin" *Pharmacological Rev. 30* (1979) 293–331; P.B. Curtis-Prior, *Prostaglandins, an Introduction to Their Biochemistry, Physiology and Pharmacology* (North-Holland, 1976).]
PG exhibit a wide variety of pharmacological properties. Of particular importance are: bronchoplasmolytic activity (treatment of acute asthma); control of gastric secretion (possible ulcer therapy); antagonism of the diuretic action of angiotensin (treatment of essential hypertension and cardiovascular disorders); initiation of ovulation, e.g. in cows, pigs and sheep; relief of pain during parturition; and finally, very small amounts of PG cause abortion.

Prostanoic acid: see Prostaglandins.

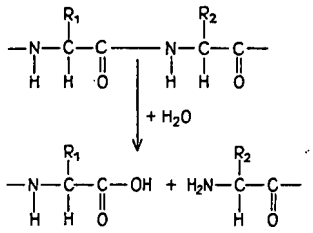
Prosthetic group: nonproteogenic, low M_r groups present in conjugated proteins. In an enzyme, a P.g. is a catalytically active group attached to the enzyme protein (apoenzyme). In the wide sense, P.g. are therefore Coenzymes (see), but many authors make a clear distinction, i.e. P.g. are covalently bound (e.g. heme, biotin, phosphopantetheine), whereas coenzymes are freely dissociable and can be removed by dialysis (e.g. pyridoxal phosphate). The flavins (FAD and FMN) of flavoproteins and flavoenzymes are usually regarded as P.g.

Protamines: a group of strongly basic, low M_r, simple (unconjugated) proteins associated with DNA in the cell nucleus. They replace the somatic Histones (see) in sperm, at least during spermogenesis. P. have been prepared from fish and bird sperm; they contain 80–85% arginine, the remaining amino acids being alanine, glycine, proline, serine and valine (or isoleucine). M_r of P. is between 4050 (clupein, 30 amino acid residues) and 4420 (iridin, 33 amino acid residues). The sequence of salmin A-I is Pro-Arg_4-Ser_3-Arg-Pro-Val-Arg_5-Pro-Arg-Val-Ser-Arg_6-Gly_2-Arg_4. Protamine-like proteins, containing 40–80 amino acid residues, have been isolated from invertebrate sperm.

Protease inhibitors: see Snake venom.

Proteases: all enzymes that catalyse the exergonic hydrolysis of peptide bonds in proteins and

peptides (Fig.). P. are divided into two groups, depending on their site of attack on the polypeptide chain:



1. *Endopeptidases (proteinases)* catalyse the hydrolysis of bonds within the peptide chain, forming variously sized cleavage peptides. They can be further subdivided into acidic, neutral and basic endopeptidases. Neutral and basic types can each be divided into Serine proteases (see) and thiol proteinases (see Thiol enzymes). Examples of animal proteinases are Pepsin (see), Rennet enzyme (see), Trypsin (see), Elastase (see), Thrombin (see), Plasmin (see) and Renin (see). For examples of plant and bacterial proteinases see Papain, Subtilisin, Bromelain. There are also yeast and fungal proteases.
2. *Exopeptidases* catalyse the hydrolytic removal of only terminal amino acids from the polypeptide chain. They can therefore be classified into *N*-terminal exopeptidases (aminopeptidases, e.g. leucine aminopeptidase) and *C*-terminal exopeptidases (e.g. carboxypeptidases A, B, Y etc.). Tri- and dipeptidases are also classified as exopeptidases.

Owing to their central importance in protein degradation and turnover, P. are widely distributed in the organism. Especially high concentrations are found in the digestive tract and in lysosomes, where they catalyse total degradation of dietary and cellular proteins, respectively, to amino acids. Most known P. are unconjugated metalloenzymes, containing or requiring a metal (e.g. zinc) as an activator or stabilizer. Self digestion (autolysis) is avoided by synthesis as inactive precursors (see Zymogens), by the presence of specific protease inhibitors, or by storage in special cell organelles called lysosomes.

Substrate specificity of P. may be high, e.g. rennet enzyme, carboxypeptidase B, enterokinase, etc., or low, e.g. pronase, pepsin and intracellular proteinases. Often P. are specific for certain amino acid residues, e.g. trypsin catalyses hydrolysis of arginyl and lysyl bonds. Pancreatic P. are among the most extensively studied of all enzymes.

Proteid: an obsolete name for a conjugated protein (see Proteins).

Protein: a naturally occurring polymer of high M_r, consisting predominantly of amino acids linked by peptide bonds. P. account for more than 50% of the organic constituents of protoplasm. By weight, they are the major component of the dry material of living organisms, and they are among the most important functional components of the living cell. All enzymes are P., and enzymes catalyse the many biochemical reactions that constitute the metabolism of the living cell; these reactions are controlled by modification of the activity and/or quantity (i.e. rate of synthesis) of appropriate enzymes. Some of the regulators and transmitters in these control processes are also P., e.g. the proteohormones (see Hormones), repressor and activator molecules (see Gene activation; Operon), and the membrane P. which determine intracellular concentrations of many enzyme substrates and products (see Active transport). Structural P. contribute to the mechanical structure of organs and tissues (e.g. elastin, collagen), or they may constitute the bulk of a natural structure (e.g. the silk of insect cocoons and spiders' webs is the protein fibroin; feathers, hair, nails and hoofs are mainly α-keratin). Contractile P. (e.g. actin and myosin of muscle, and dynein of cilia and flagella) are responsible for active movement in living organisms. Examples of storage P. are ovalbumin (from egg white), casein (from milk), gliadin (wheat seeds) and zein (maize seeds). Examples of transport P. are hemoglobin (transport of oxygen in vertebrate blood) and serum albumin (transport of fatty acids and many other substances, including hormones, drugs, etc.). P. are also involved in biochemical defense, e.g. antibodies, complement, interferons and various P. of the blood-clotting system. Certain bacterial and plant toxins are P. (see Toxic proteins). Snake venoms (see) contain various enzymes and toxic P. P. exposed on the surfaces of cells permit recognition of one cell type by another, and thus have a role in morphogenesis and the recognition of foreign tissue as such (as in graft rejection.)

The behavior and properties of P. are determined by their structures. Theoretically, there is no limit to polypeptide chain length, and all permutations and combinations of the 20 constituent amino acids are possible (see Genetic code). Further possibilities for structural variation are offered by Post-translational modification of proteins (see), attachment of non-protein prosthetic groups, and different levels of quaternary structure, so that the possible diversity of structure (and therefore function) is almost limitless.

P. in solution are neither rigid nor motionless. The bonds between the carbon atoms in the protein backbone and amino acid side chains allow considerable rotational and torsional flexibility, and the thermal motions of the individual atoms therefore produce writhing motions of the chain. These motions are almost certainly important to the enzymatic activity of the P.; and mutations which affect the flexibility of the chain could be expected to affect the kinetics of the enzyme's catalysis as well. (See Allostery.) [R. H. Pain, Nature 305 (1983) 581–582]

Classification: A logical and systematic classification of the many thousand (about 40000 in the human) P. from all living species would only be possible on the basis of their primary and tertiary structures. A simple division into enzyme P. and nonenzyme P. is meaningful and practical. A further useful classification is: simple (pure, or unconjugated) P. and conjugated P. The simple P. can be further subdivided, according to their solubility properties and molecular shape, into water-soluble globular and water-insoluble fibrous P. (Table). Alternatively, P. may be classified according to their origin, e.g. viral, bacterial, plant and animal P. Different P. within one organism may be classified as blood, milk, cerebrospinal, secretory, muscle or structural P., etc. Subcellular lo-

dnavsprotein_cypx.txt

```
>UniRef100|O34926| Cluster: Putative cytochrome P450 cypX; n=1;
            Bacillus subtilis|Rep: Putative cytochrome P450 cypX -
            Bacillus subtilis
          Length = 405

 Score =  805 bits (2080), Expect = 0.0
 Identities = 405/405 (100%), Positives = 405/405 (100%)
 Frame = +1

Query: 1    MSQSIKLFSVLSDQFQNNPYAYFSQLREEDPVHYEESIDSYFISRYHDVRYILQHPDIFT 180
            MSQSIKLFSVLSDQFQNNPYAYFSQLREEDPVHYEESIDSYFISRYHDVRYILQHPDIFT
Sbjct: 1    MSQSIKLFSVLSDQFQNNPYAYFSQLREEDPVHYEESIDSYFISRYHDVRYILQHPDIFT 60

Query: 181  TKSLVERAEPVMRGPVLAQMHGKEHSAKRRIVVRSFIGDALDHLSPLIKQNAENLLAPYL 360
            TKSLVERAEPVMRGPVLAQMHGKEHSAKRRIVVRSFIGDALDHLSPLIKQNAENLLAPYL
Sbjct: 61   TKSLVERAEPVMRGPVLAQMHGKEHSAKRRIVVRSFIGDALDHLSPLIKQNAENLLAPYL 120

Query: 361  ERGKSDLVNDFGKTFAVCVTMDMLGLDKRDHEKISEWHSGVADFITSISQSPEARAHSLW 540
            ERGKSDLVNDFGKTFAVCVTMDMLGLDKRDHEKISEWHSGVADFITSISQSPEARAHSLW
Sbjct: 121  ERGKSDLVNDFGKTFAVCVTMDMLGLDKRDHEKISEWHSGVADFITSISQSPEARAHSLW 180

Query: 541  CSEQLSQYLMPVIKERRVNPGSDLISILCTSEYEGMALSDKDILALILNVLLAATEPADK 720
            CSEQLSQYLMPVIKERRVNPGSDLISILCTSEYEGMALSDKDILALILNVLLAATEPADK
Sbjct: 181  CSEQLSQYLMPVIKERRVNPGSDLISILCTSEYEGMALSDKDILALILNVLLAATEPADK 240

Query: 721  TLALMIYHLLNNPEQMNDVLADRSLVPRAIAETLRYKPPVQLIPRQLSQDTVVGGMEIKK 900
            TLALMIYHLLNNPEQMNDVLADRSLVPRAIAETLRYKPPVQLIPRQLSQDTVVGGMEIKK
Sbjct: 241  TLALMIYHLLNNPEQMNDVLADRSLVPRAIAETLRYKPPVQLIPRQLSQDTVVGGMEIKK 300

Query: 901  DTIVFCMIGAANRDPEAFEQPDVFNIHREDLGIKSAFSGAARHLAFGSGIHNCVGAAFAK 1080
            DTIVFCMIGAANRDPEAFEQPDVFNIHREDLGIKSAFSGAARHLAFGSGIHNCVGAAFAK
Sbjct: 301  DTIVFCMIGAANRDPEAFEQPDVFNIHREDLGIKSAFSGAARHLAFGSGIHNCVGAAFAK 360

Query: 1081 NEIEIVANIVLDKMRNIRLEEDFCYAESGLYTRGPVSLLVAFDGA 1215
            NEIEIVANIVLDKMRNIRLEEDFCYAESGLYTRGPVSLLVAFDGA
Sbjct: 361  NEIEIVANIVLDKMRNIRLEEDFCYAESGLYTRGPVSLLVAFDGA 405


>UniRef100|Q65EX2| Cluster: CypX; n=1; Bacillus licheniformis ATCC
            14580|Rep: CypX - Bacillus licheniformis (strain DSM 13 /
            ATCC 14580)
          Length = 406

 Score =  648 bits (1672), Expect = 0.0
 Identities = 311/403 (77%), Positives = 363/403 (89%)
 Frame = +1

Query: 1    MSQSIKLFSVLSDQFQNNPYAYFSQLREEDPVHYEESIDSYFISRYHDVRYILQHPDIFT 180
            M+QS+K FSVLS+Q+  NPY YFS LRE DPVHYEES+DSYFISRY DVR +LQ+ D+FT
Sbjct: 1    MNQSLKTFSVLSEQYHENPYQYFSYLRESDPVHYEESLDSYFISRYQDVRRVLQNQDVFT 60

Query: 181  TKSLVERAEPVMRGPVLAQMHGKEHSAKRRIVVRSFIGDALDHLSPLIKQNAENLLAPYL 360
            TKSL +RAEPVMRGPVLAQM GKEH+AKRRIV+R FIG++LDHL+PLIK+NA+ LLAP++
Sbjct: 61   TKSLAKRAEPVMRGPVLAQMKGKEHTAKRRIVLRRFIGESLDHLTPLIKENAQRLLAPHV 120

Query: 361  ERGKSDLVNDFGKTFAVCVTMDMLGLDKRDHEKISEWHSGVADFITSISQSPEARAHSLW 540
            E+G+ DLVNDFGKTFAVCVTMD+LGLDK DH+ +  WHSGVADFITS++Q+PE R HSL
Sbjct: 121  EKGRIDLVNDFGKTFAVCVTMDILGLDKNDHQTVRNWHSGVADFITSLNQAPEDREHSLK 180

Query: 541  CSEQLSQYLMPVIKERRVNPGSDLISILCTSEYEGMALSDKDILALILNVLLAATEPADK 720
            CSEQL++YL P+I+ERR NPG DLISILCTSEYEG+A+SD+DI ALILN+LLAATEPADK
Sbjct: 181  CSEQLAEYLNPIIEERRKNPGHDLISILCTSEYEGVAMSDRDIRALILNILLAATEPADK 240
```

```
Query: 721  TLALMIYHLLNNPEQMNDVLADRSLVPRAIAETLRYKPPVQLIPRQLSQDTVVGGMEIKK 900
            TLALMIYHLL++P+QMNDVL DR+L+P+AIAETLRYKPPVQLIPRQLSQD  +GG+E+K+
Sbjct: 241  TLALMIYHLLHHPDQMNDVLEDRTLLPQAIAETLRYKPPVQLIPRQLSQDAEIGGVELKE 300

Query: 901  DTIVFCMIGAANRDPEAFEQPDVFNIHREDLGIKSAFSGAARHLAFGSGIHNCVGAAFAK 1080
             T VFCMIGAANRDPEAFE PD FNIHR DL +KSAFSGAARHLAFGSG+HNCVGA FAK
Sbjct: 301  GTTVFCMIGAANRDPEAFEDPDKFNIHRSDLEVKSAFSGAARHLAFGSGVHNCVGAGFAK 360

Query: 1081 NEIEIVANIVLDKMRNIRLEEDFCYAESGLYTRGPVSLLVAFD 1209
             EIE+VANIVLD+++NIRLEEDF Y E+GLYTRGPVSL + FD
Sbjct: 361  TEIELVANIVLDQLKNIRLEEDFIYRETGLYTRGPVSLNIRFD 403


>UniRef100|Q3F0K6| Cluster: Cytochrome p450; n=1; Bacillus
           thuringiensis serovar israelensis ATCC 35646|Rep:
           Cytochrome p450 - Bacillus thuringiensis serovar
           israelensis ATCC 35646
         Length = 406

 Score =  606 bits (1562), Expect = e-171
 Identities = 293/403 (72%), Positives = 346/403 (85%)
 Frame = +1

Query: 1    MSQSIKLFSVLSDQFQNNPYAYFSQLREEDPVHYEESIDSYFISRYHDVRYILQHPDIFT 180
            MS +I+  ++L+++FQ NPY YFS LR+ DPVHYE  IDSYFISRY DVR IL   + FT
Sbjct: 1    MSNTIQTVNILTEEFQENPYKYFSYLRQNDPVHYEPEIDSYFISRYQDVRNILNDTETFT 60

Query: 181  TKSLVERAEPVMRGPVLAQMHGKEHSAKRRIVVRSFIGDALDHLSPLIKQNAENLLAPYL 360
            TKSL ERAEPVMRGPVLAQM GKEH AKR+IV+RSF+GDAL  L PLIK+NAE+LL P+L
Sbjct: 61   TKSLAERAEPVMRGPVLAQMRGKEHVAKRKIVLRSFMGDALQKLMPLIKKNAEDLLFPHL 120

Query: 361  ERGKSDLVNDFGKTFAVCVTMDMLGLDKRDHEKISEWHSGVADFITSISQSPEARAHSLW 540
              GK DL+NDFG+TFAV VTMDM+GLDK+DH+KI EWHSGVADFITSI+Q PEA+ HSLW
Sbjct: 121  PNGKIDLINDFGRTFAVYVTMDMIGLDKKDHKKIGEWHSGVADFITSINQPPEAKKHSLW 180

Query: 541  CSEQLSQYLMPVIKERRVNPGSDLISILCTSEYEGMALSDKDILALILNVLLAATEPADK 720
            CSEQL+ YL P+IKERR+NP  DLIS LC+++YEG+A+++ DILALILN+LLAATEP DK
Sbjct: 181  CSEQLANYLEPIIKERRLNPQEDLISKLCSAKYEGIAMTNTDILALILNILLAATEPGDK 240

Query: 721  TLALMIYHLLNNPEQMNDVLADRSLVPRAIAETLRYKPPVQLIPRQLSQDTVVGGMEIKK 900
            TLAL+IY+L+N P+Q+ DVL+DRSLVP AIAETLRY PPVQLIPRQLS+DT + G+++ K
Sbjct: 241  TLALLIYNLINQPKQLQDVLSDRSLVPLAIAETLRYNPPVQLIPRQLSKDTDISGIQLSK 300

Query: 901  DTIVFCMIGAANRDPEAFEQPDVFNIHREDLGIKSAFSGAARHLAFGSGIHNCVGAAFAK 1080
             T VFCMIGAANRDP AFE+PD FNI+R DL IK AFSGAARHLAFGSGIHNCVGAAFAK
Sbjct: 301  GTTVFCMIGAANRDPNAFERPDEFNIYRPDLDIKKAFSGAARHLAFGSGIHNCVGAAFAK 360

Query: 1081 NEIEIVANIVLDKMRNIRLEEDFCYAESGLYTRGPVSLLVAFD 1209
            +EIEIV N+VLD M+NI+LEEDF Y E GLYTRGP+S+ + FD
Sbjct: 361  SEIEIVVNVVLDNMKNIKLEEDFQYVEKGLYTRGPISMPILFD 403
```

yvmc_blastx.txt

BLASTX 1.5.4-Paracel [2003-06-05]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Query= B. subtilis 168|BG14121|yvmC: 744 bp - unknown; similar to unknown proteins
(744 letters)

Database: novo2/protein/UniRef100
4,953,717 sequences; 1,776,136,764 total letters

Searching.....done

Sequences producing significant alignments:	Score (bits)	E Value
UniRef100|O34351| Cluster: YvmC; n=1; Bacillus subtilis|Rep: Yvm...	494	e-138
UniRef100|Q65EX3| Cluster: YvmC; n=1; Bacillus licheniformis ATC...	345	1e-93
UniRef100|Q3F0K7| Cluster: Putative uncharacterized protein; n=1...	277	4e-73

>UniRef100|O34351| Cluster: YvmC; n=1; Bacillus subtilis|Rep: YvmC - Bacillus subtilis
Length = 248

Score = 494 bits (1273), Expect = e-138
Identities = 248/248 (100%), Positives = 248/248 (100%)
Frame = +1

```
Query: 1   MTGMVTERRSVHFIAEALTENCREIFERRRHVLVGISPFNSRFSEDYIYRLIGWAKAQFK 180
           MTGMVTERRSVHFIAEALTENCREIFERRRHVLVGISPFNSRFSEDYIYRLIGWAKAQFK
Sbjct: 1   MTGMVTERRSVHFIAEALTENCREIFERRRHVLVGISPFNSRFSEDYIYRLIGWAKAQFK 60

Query: 181 SVSVLLAGHEAANLLEALGTPRGKAERKVRKEVSRNRRFAERALVAHGGDPKAIHTFSDF 360
           SVSVLLAGHEAANLLEALGTPRGKAERKVRKEVSRNRRFAERALVAHGGDPKAIHTFSDF
Sbjct: 61  SVSVLLAGHEAANLLEALGTPRGKAERKVRKEVSRNRRFAERALVAHGGDPKAIHTFSDF 120

Query: 361 IDNKAYQLLRQEVEHAFFEQPHFRHACLDMSREAIIGRARGVSLMMEEVSEDMLNLAVEY 540
           IDNKAYQLLRQEVEHAFFEQPHFRHACLDMSREAIIGRARGVSLMMEEVSEDMLNLAVEY
Sbjct: 121 IDNKAYQLLRQEVEHAFFEQPHFRHACLDMSREAIIGRARGVSLMMEEVSEDMLNLAVEY 180

Query: 541 VIAELPFFIGAPDILEVEETLLAYHRPWKLGEKISNHEFSICMRPNQGYLIVQEMAQMLS 720
           VIAELPFFIGAPDILEVEETLLAYHRPWKLGEKISNHEFSICMRPNQGYLIVQEMAQMLS
Sbjct: 181 VIAELPFFIGAPDILEVEETLLAYHRPWKLGEKISNHEFSICMRPNQGYLIVQEMAQMLS 240

Query: 721 EKRITSEG 744
           EKRITSEG
Sbjct: 241 EKRITSEG 248
```

>UniRef100|Q65EX3| Cluster: YvmC; n=1; Bacillus licheniformis ATCC 14580|Rep: YvmC - Bacillus licheniformis (strain DSM 13 / ATCC 14580)
Length = 249

Score = 345 bits (886), Expect = 1e-93
Identities = 174/247 (70%), Positives = 201/247 (80%)
Frame = +1

```
Query: 1   MTGMVTERRSVHFIAEALTENCREIFERRRHVLVGISPFNSRFSEDYIYRLIGWAKAQFK 180
```

```
            MT ++ E +    F  E LT+NC EI +RRRHVLVGISPFNSRFSEDYI+RLI WA  +F+
Sbjct: 1    MTELIMESKHQLFKTETLTQNCNEILKRRRHVLVGISPFNSRFSEDYIHRLIAWAVREFQ 60

Query: 181 SVSVLLAGHEAANLLEALGTPRGKAERKVRKEVSRNRRFAERALVAHGGDPKAIHTFSDF 360
            SVSVLLAG EAANLLEALGTP GKAERKVRKEVSRNRRFAE+AL AHGG+P+ IHTFSDF
Sbjct: 61   SVSVLLAGKEAANLLEALGTPHGKAERKVRKEVSRNRRFAEKALEAHGGNPEDIHTFSDF 120

Query: 361 IDNKAYQLLRQEVEHAFFEQPHFRHACLDMSREAIIGRARGVSLMMEEVSEDMLNLAVEY 540
             +  AY+ LR EVE AFF+Q HFR+ACL+MS  AI+GRARG  + + EVS DML LAVEY
Sbjct: 121  ANQTAYRNLRMEVEAAFFDQTHFRNACLEMSHAAILGRARGTRMDVVEVSADMLELAVEY 180

Query: 541 VIAELPFFIGAPDILEVEETLLAYHRPWKLGEKISNHEFSICMRPNQGYLIVQEMAQMLS 720
            VIAELPFFI APDIL VEETLLAYHRPWKLGE+IS +EF++ MRPNQGYL+V E  + +
Sbjct: 181  VIAELPFFIAAPDILGVEETLLAYHRPWKLGEQISRNEFAVKMRPNQGYLMVSEADERVE 240

Query: 721 EKRITSE 741
             K +  E
Sbjct: 241  SKSMQEE 247


>UniRef100|Q3F0K7| Cluster: Putative uncharacterized protein; n=1;
           Bacillus thuringiensis serovar israelensis ATCC
           35646|Rep: Putative uncharacterized protein - Bacillus
           thuringiensis serovar israelensis ATCC 35646
          Length = 238

 Score =  277 bits (709), Expect = 4e-73
 Identities = 137/238 (57%), Positives = 177/238 (73%)
 Frame = +1

Query: 1    MTGMVTERRSVHFIAEALTENCREIFERRRHVLVGISPFNSRFSEDYIYRLIGWAKAQFK 180
            MT  +  R    F ++ L+ NC EI +R +H ++GISPFNSRFS++YI RLI WA   F
Sbjct: 1    MTNAIAVRNVRKFSSQPLSTNCAEILKRSKHAIIGISPFNSRFSDEYINRLIEWALHTFD 60

Query: 181 SVSVLLAGHEAANLLEALGTPRGKAERKVRKEVSRNRRFAERALVAHGGDPKAIHTFSDF 360
             VSVLLAG EAANLLEALG  + K + K+ K++SRNRR AE+AL  HGG+  AIHTFSDF
Sbjct: 61   DVSVLLAGKEAANLLEALGH-QSKRKEKLGKKISRNRRSAEKALKEHGGNVNAIHTFSDF 119

Query: 361 IDNKAYQLLRQEVEHAFFEQPHFRHACLDMSREAIIGRARGVSLMMEEVSEDMLNLAVEY 540
             DN AY  +R E EH F  +  FR+ACL+MS  AI+GRARG ++ ++++S DMLN+AVEY
Sbjct: 120  NDNNAYSCMRAEAEHIFLSETVFRNACLEMSHAAILGRARGTNIDIDQISNDMLNIAVEY 179

Query: 541 VIAELPFFIGAPDILEVEETLLAYHRPWKLGEKISNHEFSICMRPNQGYLIVQEMAQM 714
            VIAELPFFIG  +IL  +E +L YH+PW+LGE+I  ++FSI M+PNQGYL+VQEM  +
Sbjct: 180  VIAELPFFIGGAEILGTQEAVLIYHKPWELGEQIVRNDFSIRMKPNQGYLMVQEMENL 237


  Database: novo2/protein/UniRef100
    Posted date:  Aug 21, 2007  2:45 AM
  Number of letters in database: 1,776,136,764
  Number of sequences in database:  4,953,717

Lambda     K      H
   0.318    0.135    0.401

Gapped
Lambda     K      H
   0.267   0.0410    0.140


Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
```

Number of HSP's successfully gapped in prelim test: 0
Number of HSP's that attempted gapping in prelim test: 0
length of database: 1,776,136,764
effective HSP length: 130
effective length of database: 1,132,153,554
effective search space used: 132461965818
frameshift window, decay const: 50, 0.1
T: 12
A: 40
X1: 16 (7.3 bits)
X2: 38 (14.6 bits)
X3: 64 (24.7 bits)
S1: 41 (21.7 bits)